



Workshop on acceleration of the validation and regulatory acceptance of alternative methods and implementation of testing strategies



A.H. Piersma^{a,b,*}, T. Burgdorf^c, K. Louekari^d, B. Desprez^e, R. Taalman^e, R. Landsiedel^f,
J. Barroso^g, V. Rogiers^h, C. Eskesⁱ, M. Oelgeschläger^c, M. Whelan^g, A. Braeuning^k,
A.M. Vinggaard^j, A. Kienhuis^a, J. van Benthem^a, J. Ezendam^a

^a National Institute for Public Health and the Environment (RIVM), Center for Health Protection, Bilthoven, Netherlands

^b Institute for Risk Assessment Sciences, Utrecht University, Netherlands

^c German Centre for The Protection of Laboratory Animals (Bf3R), German Federal Institute for Risk Assessment, Berlin, Germany

^d European Chemicals Agency, Helsinki, Finland

^e Cosmetics Europe, Brussels, Belgium

^f BASF SE, Ludwigshafen, Germany

^g European Commission, Joint Research Centre (JRC), Ispra, Italy

^h Vrije Universiteit Brussel, Brussels, Belgium

ⁱ SeCAM, Magliaso, Switzerland

^j National Food Institute, Technical University of Denmark, Kgs. Lyngby, Denmark

^k German Federal Institute for Risk Assessment, Berlin, Germany

A B S T R A C T

This report describes the proceedings of the BfR-RIVM workshop on validation of alternative methods which was held 23 and 24 March 2017 in Berlin, Germany. Stakeholders from governmental agencies, regulatory authorities, universities, industry and the OECD were invited to discuss current problems concerning the regulatory acceptance and implementation of alternative test methods and testing strategies, with the aim to develop feasible solutions. Classical validation of alternative methods usually involves one to one comparison with the gold standard animal study. This approach suffers from the reductionist nature of an alternative test as compared to the animal study as well as from the animal study being considered as the gold standard. Modern approaches combine individual alternatives into testing strategies, for which integrated and defined approaches are emerging at OECD. Furthermore, progress in mechanistic toxicology, e.g. through the adverse outcome pathway approach, and in computational systems toxicology allows integration of alternative test battery results into toxicity predictions that are more fine-tuned to the human situation. The road towards transition to a mechanistically-based human-focused hazard and risk assessment of chemicals requires an open mind towards stepping away from the animal study as the gold standard and defining human biologically based regulatory requirements for human hazard and risk assessment.

1. General introduction

Tanja Burgdorf (BfR)

Current validation procedures assess the reliability and relevance of individual test methods. These procedures are essential to investigate the technical validity and transferability of a method and are important in the process of regulatory acceptance, e.g. as an Organization for Economic Cooperation and Development (OECD) test guideline. However, the validation process, as described in the OECD Guidance document 34 (OECD, 2005), requires significant time and financial resources demanding for a target-oriented and efficient selection and

evaluation process.

Non-animal toxicological methods (e.g. *in vitro* assays, *in silico* models, – omics data) become increasingly available. Single alternative test methods can generally not replace *in vivo* test methods due to the inherent biological complexity of adverse effects. Hence different testing methods need to be combined into integrated or sequential testing strategies in order to allow reliable prediction of adverse effects *in vivo* (Piersma et al., 2014). However, up to now, there is no overarching strategy how to assess the predictivity and reproducibility of testing strategies for a specific regulatory need. The absence of such a procedure is an important hurdle in the process towards regulatory

* Corresponding author at: RIVM, P.O. Box 1, 3720 BA Bilthoven, Netherlands.
E-mail address: aldert.piersma@rivm.nl (A.H. Piersma).

acceptance and implementation of alternative methods. Particularly, in contrast to OECD test guidelines covering single test methods, testing strategies are not currently covered by the OECD Council Decision on Mutual Acceptance of Data (MAD).

In recent years the concept of Integrated Approaches on Testing and Assessment (IATA) evolved at the OECD level and combines different methodological approaches in an iterative manner. In 2014 an IATA for skin corrosion/irritation was published as a first example and an IATA for eye damage/irritation was published recently (OECD, 2014a; OECD, 2017a). The concept of Adverse Outcome Pathways (AOPs) was developed to organize the current mechanistic understanding and toxicological data on different biological levels (Ankley et al., 2010). AOPs can also represent a framework to develop non-animal alternatives to *in vivo* testing as it has been shown for the AOP on skin sensitization. This AOP triggered the development of key-event based predictive alternative methods and facilitated their regulatory acceptance following publication of the corresponding OECD test guidelines (OECD, 2012a; OECD, 2012b; OECD, 2015a; OECD, 2015b; OECD, 2016a). AOP-based IATAs are increasingly discussed and OECD guidance has been developed accordingly (Tollefsen et al., 2014; Edwards et al., 2015; OECD, 2016b). IATAs were intended to be flexible and always include a degree of expert judgment to meet specific regulatory needs. Some elements of an IATA, however, can be standardized and interpreted according to well-characterized and transparent prediction models. These so-called Defined Approaches that consist of a testing strategy and a fixed data interpretation procedure (DIP) are currently discussed as a future tool in chemical safety assessment. In a proof-of-concept study 12 different Defined Approaches (DA) for skin sensitization have been established and described according to OECD guidance (OECD, 2016c). The rapid congruent progress in the development of AOPs and novel, innovative alternative methods will further increase the need to integrate information from multiple sources into testing strategies (or defined approaches). The question arises how these test strategies can be harmonized and standardized and if and how they can be integrated into legal frameworks such as the OECD test guidelines programme to be covered by the MAD principle and to achieve legal certainty which is a prerequisite for their regulatory applicability. Therefore, a discussion started how these new paradigms can be integrated in the safety assessment of chemicals (Burden et al., 2015; Ramirez et al., 2015).

To foster this dialogue, the German Federal Institute for Risk Assessment (BfR) in cooperation with the Dutch National Institute for Public Health and the Environment (RIVM), organized a joint workshop on 23 and 24 March 2017 in Berlin. International experts from governmental agencies, regulatory authorities, universities, industry and the OECD were invited to discuss current problems concerning the regulatory acceptance and implementation of alternative test methods and testing strategies to develop feasible solutions.

In a series of introductory talks, drivers and barriers of the process of regulatory acceptance of test strategies were presented. Experts from RIVM, the European Chemicals Agency (ECHA), the EU Joint Research Center in Ispra, Italy (JRC), Scientific Committee on Consumer Safety (SCCS), Cosmetics Europe and BASF described their point of view which provided the basis for a focused discussion in breakout groups during which three main questions were addressed. The first group discussed how the quality of alternative methods, in particular as a part of a testing strategy, can be improved, including the question for criteria that need to be defined prior to validation of a single test to assess its added value and position in a test strategy. The second group considered the important question how the validation process can be accelerated. Not only the question how the transfer from test method development into validation can be promoted was discussed but also if new methods should always be validated as part of testing strategies. The third group addressed the questions how a testing strategy has to be designed and validated in such a way that it satisfies the regulatory needs and how testing strategies could be on equal footing with individual test guidelines.

2. Lecture summaries

2.1. Mechanistic and computational toxicology using ontologies, and their consequences for validation of animal-free methods

Aldert Piersma (RIVM)

Human risk of chemical exposures is classically assessed based on toxicity studies in experimental animals. This practice is showing its limitations. Animal studies are not the perfect predictor for human hazard and risk, and ethical issues around animal experimentation have increasingly challenged the current hazard and risk assessment practice. Both these issues have stimulated the development and application of animal-free alternative methods for chemical hazard and risk assessment. Such alternative methods can be fine-tuned to the human situation since they can be developed from biological materials of human origin and can provide mechanistic information that can help to better understand hazard as opposed to the black box system of studying adverse health effects in the intact animal. The validation of alternative methods has classically been done by one to one comparisons with the gold standard animal study based on a limited number of reference chemicals. Predictivity was limited by the reductionist nature of the alternative test as compared to the animal study, and lacked the comparison with the species of interest, which is man. This practice has hampered implementation of animal-free methods in regulatory hazard and risk assessment.

Mechanistic toxicology has received more and more attention since the publication of the US National Academy of Sciences report on Toxicity Testing in the 21st Century (USA NNAoS, 2007). OECD has adopted the Adverse Outcome Pathway (AOP) approach and defined Integrated Approaches to testing and Assessment (IATA), and Defined Approaches (OECD, 2012b; OECD, 2016b; OECD, 2016c). AOPs have helped identify pathways from initiating events to adverse outcomes *via* key events in physiology. In order to understand toxicity at the level of the intact human, AOPs should be integrated to a network of pathways, in which key event relationships are described in quantitative terms. This allows the design of a systems toxicology computerized model that integrates all information and predicts toxicity at the apical level. Based on the model, critical key events that are rate-limiting for the occurrence of adverse effects can be identified. These key events can be monitored in dedicated molecular and cellular assays, including organ-on-a-chip methodologies. The outcomes of this testing battery can then be fed into the model to predict adverse health effects. The virtual embryo project at US EPA has successfully shown the feasibility of this approach for selected aspects of embryogenesis (Hutson et al., 2017; Leung et al., 2016; Kleinstreuer et al., 2013).

The information at the basis of a computational systems toxicology model can be collected and structured in an ontology. Ontologies offer the matrix for mapping subjects and their interrelationships, and can be applied to physiological parameters and their quantitative relationships. The ontology does not need to be comprehensive as to physiology, but should cover all relevant toxicity pathways, which likely represent a fraction of the physiological landscape. Current extensive knowledge about basic biology, chemistry as well as toxicology offers a wealth of data to populate the ontology. As an example, molecular mechanisms in developmental biology are rapidly being elucidated. Chemical structure information allows prediction of the developmental toxicity potential of chemicals (Wu et al., 2013). The extensive ToxCast database contains a plethora of information on the molecular and cellular effects of thousands of chemicals in hundreds of assays (Richard et al., 2016). The crux is in describing the toxicity pathway ontology to the level of detail necessary for reliable toxicity prediction. One important pathway in prenatal development is the retinoic acid homeostatic system, that is crucial for vertebrate embryogenesis (Campo-Paysaa et al., 2008). A preliminary AOP framework has been published, and several *in vitro* assays have shown regulation of the framework after chemical exposures, such as rat whole embryo culture, the zebrafish

embryo test, and the mouse and human embryonic stem cell tests (Tonk et al., 2015). The time is right for investing in the ontology approach, feeding into a computational systems toxicology approach fine-tuned to man, avoiding the detour of the animal experiment.

Given that the systems toxicology model needs to be fed by results from combinations of *in vitro* assays, this affects the way these assays should be validated. Biological domain, technical performance and chemical applicability domain remain important characteristics of individual tests. The biological domain can be validated by testing a limited number of known positive and negative exposures in the system. However, as to predictivity, an individual assay with a necessarily limited biological domain will by definition not be able to detect all toxicants. It is the predictive capacity of the combination of assays that matters for the reliability of the system as a whole. Therefore, rather than testing a series of chemicals, it is more worthwhile to assess whether the biological domain of the combination of tests sufficiently covers all toxicity pathways that need to be assessed. The current expansion of molecular and mechanistic toxicology, *in silico* methods based on chemical structure and properties, and *in vitro* animal-free methodologies pave the way for an integrated animal-free approach for dedicated human hazard and risk assessment of chemicals.

2.2. Development and regulatory approval of alternative methods

Kimmo Louekari (ECHA)

The European Regulation on Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) requires that the *in vivo* tests be only made as last resort. The European Chemicals Agency (ECHA) therefore actively contributes to the development of new guidelines and testing strategies in collaboration with OECD and the European Union Reference Laboratory for alternatives to animal testing (EURL-ECVAM). At present *in vitro* test methods and IATAs are available for all lower tier human health endpoints, *i.e.* skin and eye irritation, corrosion, and skin sensitization. Consequently, REACH information requirements and ECHA Guidance have been updated. The next goal is the development of alternatives for the “higher tier” studies, repeated dose toxicity studies as well as developmental and reproductive toxicity studies.

To examine whether development of alternative methods can be accelerated, the related processes in OECD and in ECHA were examined. Before these processes start, test methods need to be prioritized for validation. Steps related to validation take totally 3.5–8 years. The most time consuming part is the actual validation study. Some aspects of that are covered elsewhere in this article. EURL ECVAM has introduced a novel step in this process: it has founded so-called PARERE network (Preliminary Assessment of REgulatory RElevance). PARERE consists of a group of regulators who give advice on the regulatory relevance of new test submissions and strategy documents. PARERE consultations aim to ensure that the test methods entering to validation are such that they will be used in a regulatory work. This implies that a test method is likely to provide reliable and predictive data being required under relevant European legislations, *e.g.* for industrial chemicals. Indeed one possibility to accelerate the targeted development of new alternative methods is that ECHA and other regulatory agencies proactively inform the scientific community and validation bodies of the needs for alternative test methods. That will enable efficient use of the limited resources of scientific validation.

Table 1 summarises the process of guideline development in OECD. Totally this takes usually 2–5.5 years but the process may be longer, when there are scientific controversies or disagreement on the actual regulatory need of a new test guideline. Therefore, in case a test guideline development is prolonged, one OECD member country should actively propose solutions and agreeable compromises.

In Table 2, the recognition and the first steps of regulatory use of the new test methods are described. This work usually takes 1.5–5 years. In these steps, ECHA and the European Commission have an active role in analysing and communicating the correct regulatory use of test

guidelines.

A test guideline or an alternative method is a useful regulatory tool when it: (OECD, 2005) produces such data on toxicity that correspond (at least partly) with a specific information requirement, (Piersma et al., 2014) enables classification and labelling, (OECD, 2014a) enables risk assessment, which requires that the test result provide quantitative dose descriptor, to be used in setting a DNEL (Derived No Effect Level), (OECD, 2017a) can contribute to an IATA, *i.e.* that it covers certain parts of an adverse outcome pathway, with a validated *in vitro* or *in vivo* method. Successful validation, predictivity and biological relevance of the test methods are obvious prerequisites of the regulatory recognition of new test methods.

There are possibilities to accelerate the regulatory recognition and use of new test methods and alternative data. REACH Committees could deal the precedence case early and thereby communicate the acceptability of various types of new data. In fact only the discussion and decision making in REACH Committees will show in practice how the new tests can be used by the regulators, and Committee discussions could be seen as “a test of acceptability”. Secondly, ECHA could use different channels/media to efficiently inform the industry of the new methods and alternatives. In particular cases, ECHA could even “publish” validated test methods before OECD approval, to encourage their use.

2.3. Testing strategies in modern toxicology: a journey from validation to regulatory acceptance

Bertrand Desprez and Rob Taalman (Cosmetics Europe)

Current validation processes are inherently complex and contain various sources of uncertainty. Most validation bodies have a stepwise workflow. Whilst this is scientifically justified, the multiplicity of these steps as well as related and repeated clarifications and sub-evaluations are occasions of potential misunderstandings between the submitter and evaluators. This may partly explain why the pace of the validation process is relatively slow. Complexity is well reflected in the example of the skin irritation and corrosion endpoint. Not less than 24 statements, documents or guidelines were published between 1998 and 2014 when an Integrated Approach for Testing and Assessment (IATA) was finally accepted in the regulatory arena. The OECD Guidance Document No. 34 defines 7 key steps (*e.g.*, reproducibility, predictivity) of a modular approach that EURL ECVAM implements (OECD, 2005). This approach has many advantages, it is clear (target values), transparent (available to all), scientifically sound (good predictivity should be reproducible), and already demonstrated its usefulness. It may benefit from further refinement, for instance, by extending its applicability beyond individual methods to combined methods/testing strategies. Issuing formal statements would certainly provide clarity here.

The validation process has several sources of uncertainty, pertaining to scientific and legal issues, to individual *versus* combined methods, and to the gold standard animal data. From the scientific perspective, the use of single absolute target values does not reflect statistical reality. For instance, the simple use of confidence intervals, rather than target values of reproducibility or sensitivity, would be a step towards more probabilistic validations. A recent opinion of the External Scientific Advisory Committee to ECVAM (ESAC) on the use of Performance Standards to evaluate test methods similar to validated reference methods, although limited to similar tests methods (so-called “me-too” tests), provides some indications on these types of approaches (Committee EESA, 2016). That would reflect the inherent variability in assays to be included in testing strategies. Legal uncertainties remain as to which types of international agreements formally ratify the regulatory acceptance. Whilst legal clarity exists in OECD test guidelines on individual assays that are covered by Mutual Acceptance of Data (MAD), IATAs are so far described in guidance documents (GDs), and are thus not covered by MAD. Regarding combined tests, there is uncertainty on cut-off values in prediction models. Are cut-offs of

Table 1
Development of new test guidelines and Guidance Documents in OECD.

Step description	Duration
Submission of a project proposal (SPSF) to OECD, mostly by OECD member countries (MCs).	1–2 months
OECD MCs comment on the SPSF. Comments on potential regulatory use and relevance are crucial.	1–2 months
OECD Working Group of the National Coordinators for the Test Guidelines Programme (WNT) approves or rejects the test guideline project.	1 week
Draft test guideline is prepared by an OECD expert group: e.g. sensitivity, specificity, domain, limitations, and method optimisation are addressed.	18 months-up to about 5 years
OECD MCs and EU Agencies comment on the draft TGs and Guidance Documents.	2–3 months
OECD WNT discusses the draft test guidelines, approves it or sends it back to the expert group.	1 week

Table 2
Recognition and start of the regulatory use of new test guidelines under REACH.

Step description	Duration
Inclusion of the approved OECD TGs in EU Test Method Regulation, by Commission (not absolutely necessary)	1 year-10 years
Inclusion of the new method in ECHA Guidance. ECHA Guidance is authoritative in specifying the regulatory scope and limits of the use of new test guidelines.	1–2 years
Update of the ECHA test method web site or inclusion in case the TM endpoint was not there.	1–3 months
First cases dealt with in ECHAs Committees, MSCA and RAC	6 month-2 years

individual tests still fit for purpose? Cut-offs when tests are combined should be adapted, but, to date, this is not appropriately covered in GDs. As to animal data as the gold standard, in topical toxicology, there are publications showing shortcomings of the traditional animal tests. The same weight given to occurrence of an eye effect irrespective of its frequency leads to unjustified classification in the most severe category. In skin sensitization, the Local Lymph Node Assay (LLNA) used, as a reference for validation, does not have a perfect accuracy rate (Hoffmann, 2015). The question of appropriate gold standards in validation of alternative approaches is of utmost importance.

As methods are progressively being combined in testing strategies, the question arises how this affects the validation process. Performance of combined tests can be much different from that of individual tests within the combination. For example two independent test methods with sensitivity/specificity of 60%/80% and 75%/90%, respectively, would have an overall performance, in a top-down approach, of 90%/72% (Fig. 1). The overall sensitivity is increased in comparison to individual methods since a positive prediction has 2 occasions to occur (true positive with the first method OR false negative in the first method then caught up as true positive in the second one). In parallel, the overall specificity decreases in comparison to individual methods, since a negative prediction requires that the two methods provide both a negative prediction.

Whereas each test method may not be individually validated and accepted, their sensitivity in combination should be sufficient to be accepted by regulatory authorities, for predictions in the frame of safety evaluations of chemicals. Guidance needs to be developed. Easing the validation process of combined test methods requires to systematically address the uncertainties described above. This should happen at an early stage of the validation process, and covered in an additional module, expanding the existing modular approach. AOPs, IATAs, but also exposure and kinetic modelling (ADME) tools are gaining importance in novel safety assessment approaches. More governance is needed on how to achieve their mutual and harmonized acceptance, as opposed to the well-defined acceptance of individual methods in test guidelines. Pending key questions regarding mechanistically-related considerations and their acceptance include: (i) what is the amount of MIEs/KEs from a known AOP that a testing strategy should be covering?; (ii) how can we achieve further development and application of a network of AOPs or even mode of action-based ontologies?; (iii) how can one waive binary predictions in this context and allow for

quantitative characterisation?

OECD and EURL ECVAM are good fora to address all the above topics. Current validation processes historically allowed the validation of individual alternative test methods. There is still room left for adapting the validation process to the 21st century risk assessment, and here are 5 ideas on how to do it: (OECD, 2005) establish clear timelines between all steps of the validation workflow; (Piersma et al., 2014) find flexible ways, but still formal ones, to strengthen cooperation and discussion between test developers and validation bodies, along the process. Case studies can be used as mechanisms to bring all stakeholders together; (OECD, 2014a) systematically include uncertainty considerations, sufficiently ahead of regulatory acceptance for candidate integrated approaches; (OECD, 2017a) clarify how to achieve mutual acceptance of integrated approaches or test strategies, and whether individual validation of each test method forming these approaches can be waived; (Ankley et al., 2010) reach consensus on the value of gold standards, since animal approaches as reference have been shown to be uncertain.

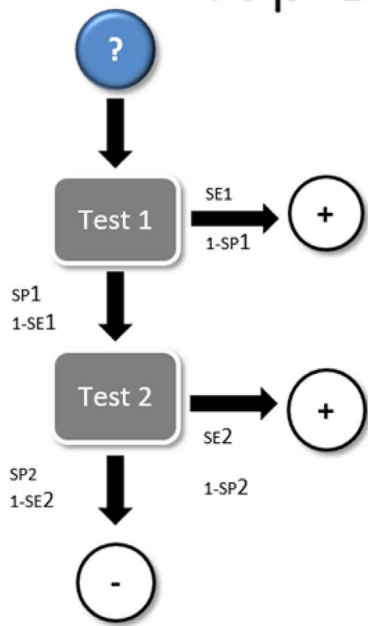
2.4. Alternative methods in the chemical industry: hurdles to regulatory use and 12 proposals to overcome them

Robert Landsiedel (BASF)

Alternative methods (and animal methods alike) have to fulfill certain requirements to be used in the chemical industry. Fig. 2 is illustrating the main requisites and prerequisites, and Fig. 3 gives an example of an administrative process for validation in the EU and adoption of an OECD test guideline. In the following, hurdles to achieve regulatory acceptance and eventually regulatory use of a new method are discussed. Most toxicological studies are a legal obligation. The REACH regulation (Commission E. Regulation (EC), 2006) defines toxicological standard information requirements in order to facilitate risk assessment and classification and labelling (Classification, Labelling & Packaging, CLP; (Commission E. Regulation (EC), 2008)). REACH Annex VII demands information on skin corrosion/irritation, eye damage/irritation, skin sensitization, mutagenicity and acute systemic toxicity for all substances with a production volume above one ton *per annum*. Often, the information requirements refer to an animal method or the regulation, like CLP, was designed with animal studies and the data they are providing in mind. Currently work is undergoing to redesign regulations, like the United Nations Globally Harmonized System (UN/GHS) or REACH/CLP, taking into account results of non-animal methods. The question arises if regulations according to adverse outcomes (“endpoints”) are the only option to enable risk assessments or if they could – at least in part – be replaced or amended by information on earlier key events and biokinetic information.

In any case, new methods to provide data on the underlying AOPs will be needed. For a global and straightforward regulatory acceptance of the results, methods need to be validated and accepted by regulatory authorities. In order to facilitate the regulatory use of a greater number of new methods addressing different key events (KE) of an AOP, the validation process and regulatory acceptance needs to be simplified and sped up. So far only few information requirements of REACH can be completely fulfilled by alternative methods (*id est.*: skin- and eye

Top-Down (TD) approach



Test1 identification of + chemicals

Test2 identification of + or - chemicals

Test1	SE1	0.60	1-SE1	0.40
	SP1	0.80	1-SP1	0.20
Test2	SE2	0.75	1-SE2	0.25
	SP2	0.90	1-SP2	0.10

		In Vivo +	In Vivo -
Top-down approach +	Test1 + OR Test1 - AND Test2 +	n1 SE1 + n1 (1-SE1).SE2	n2 (1-SP1) + n2 SP1.(1-SP2)
Top-down approach -	Test1 - AND Test2 -	n1.(1-SE1).(1-SE2)	n2.SP1.SP2
		n1	n2

Performance of the **testing strategy**

Sensitivity = 0.900 = 90.0%

Specificity = 0.720 = 72.0%

Fig. 1. Example of individual *versus* combined performance of two tests, showing enhanced sensitivity and reduced specificity of the combination as compared to individual tests.

irritation and mutagenicity; the skin sensitization hazard can be tested by alternative methods, but potency sub-classification still requires animal testing). Obviously, method development is still needed to address the majority of toxicological endpoints, which are currently addressed by animal methods. The validation and OECD TG adoption are

started by application of the test method developer *via* its member country. There is little guidance or prioritization on which methods should be developed. This leaves the OECD waiting for methods (and it is a matter of chance to have fitting methods to address an AOP adequately) and it leaves the test developer with the risk of investing in a

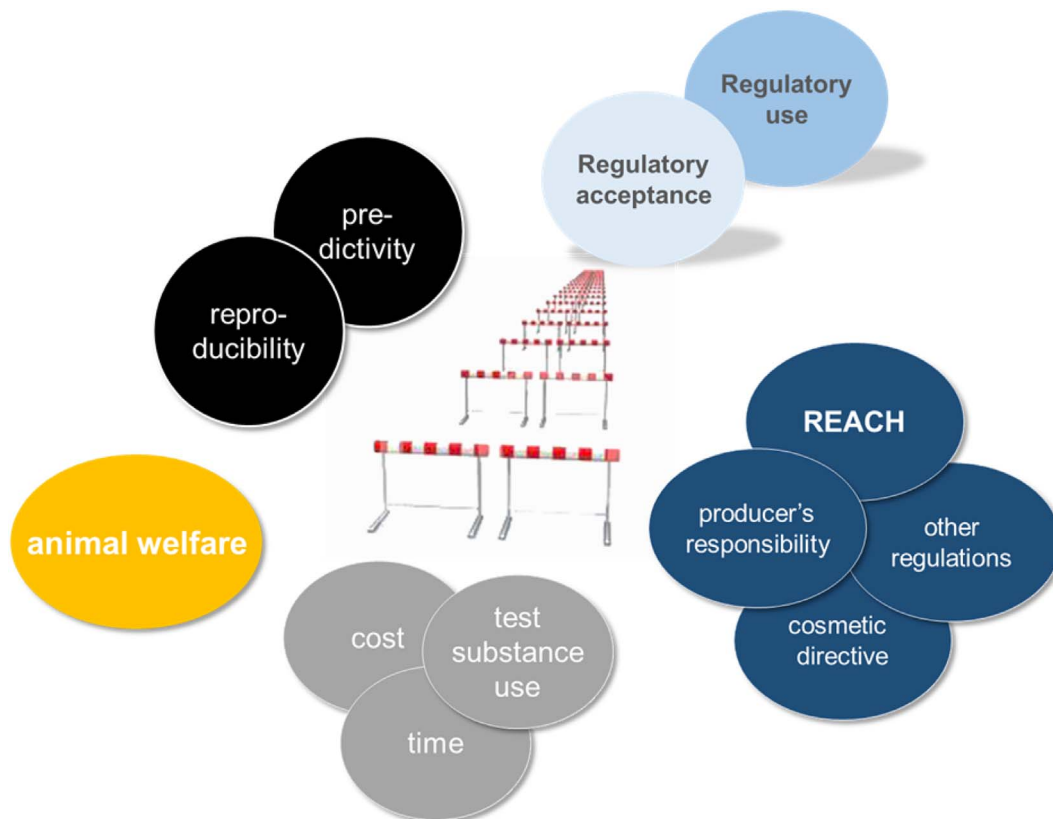


Fig. 2. Requisites and prerequisites (“hurdles”) of regulatory use of alternative methods (and animal methods alike) (Figure adapted from Landsiedel et al., 2013).

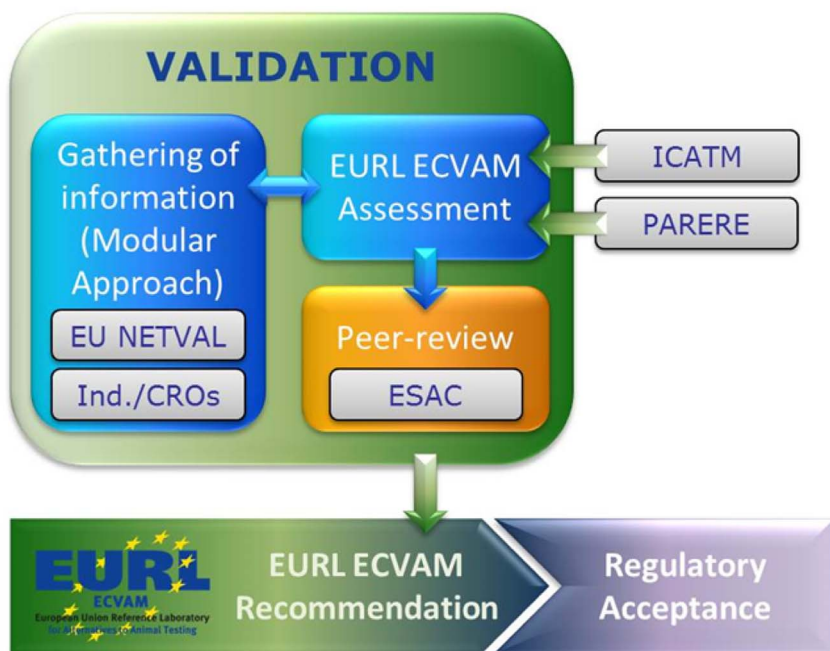
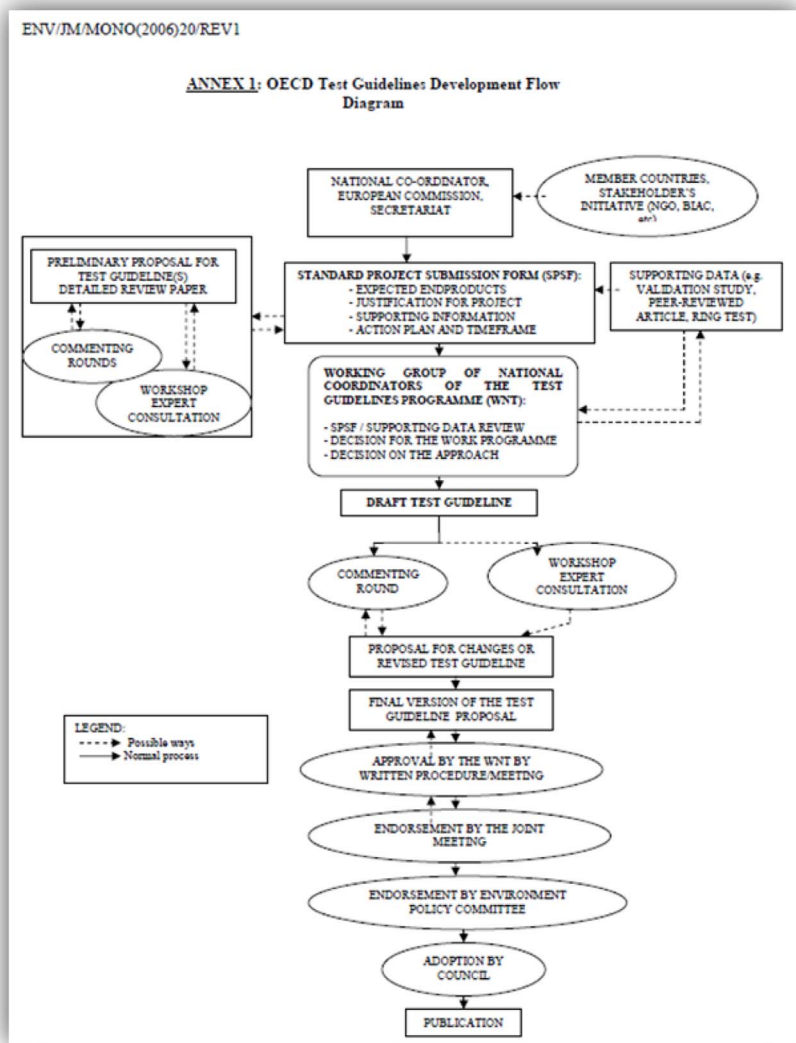


Fig. 3. Administrative process of new test method validation in the EU (upper) and adoption of OECD test guideline (lower). (EURL ECVAM's Validation Process; OECD ENV/JM/MONO (2006)20. EU NETVAL – European Union Network of Laboratories for the Validation of Alternative Methods (<https://eurl-ecvam.jrc.ec.europa.eu/eu-netval>); Ind./CROs – Industry and/or contract research organizations; ESAC - EURL ECVAM Scientific Advisory Committee (<https://eurl-ecvam.jrc.ec.europa.eu/about-ecvam/scientific-advice-stakeholders-networks/ecvam-scientific-advisory-committee-esac>); ICATM – International Cooperation on Alternative Test Methods (<https://eurl-ecvam.jrc.ec.europa.eu/about-ecvam/networks-and-collaborations/collaboration-with-icatm>); PARERE – EURL ECVAM's Network for Preliminary Assessment of Regulatory Relevance (<https://eurl-ecvam.jrc.ec.europa.eu/about-ecvam/scientific-advice-stakeholders-networks/parere>)).



method which is *ex post* not prioritized in or even not at all admitted to the OECD Test Guideline (TG) program. Hence, testing strategies according to AOPs and the requested performance of the respective methods should be pre-defined and method development should be prioritized accordingly.

In the process of TG adoption, intellectual property restrictions and limitations of the supply of biological materials are deliberated. There have, however, been cases of missing technical equipment (e.g. the opacimeter used for the bovine corneal opacity and permeability test, OECD TG no. 437; (Schrage et al., 2011)). Moreover, it is unclear, how many (proprietary) methods are desirable to provide the same information, e.g. there are currently two reconstructed tissue models to test for eye irritation properties (OECD TG 492), four different models for skin irritation and corrosion (OECD TG 431 and 439 (OECD, 2015c; OECD, 2016d), and three variants of the local lymph node assay (OECD TG 429, 442A and B) (OECD, 2010a; OECD, 2010b; OECD, 2010c) but also methods not adopted by OECD (Basketter et al., 2012; Kolle et al., 2012). It seems obvious, that more than one proprietary method should be adopted in the respective TG in order to avoid monopolies but resources may be too scarce to adopt as many as one hundred methods. This would, however, put the developer of the hundredth method at a competitive disadvantage, and prior information on whether another method will still be considered, would have saved resources (and disappointments).

Once a new method has been validated and accepted by regulatory authorities, it is generally considered “good forever”. Methods may, however, turn out to be of less or actually very limited value after they have been used for a while, e.g. the setting of starting doses for acute oral toxicity testing by cytotoxicity testing in mouse fibroblasts was shown to be ineffective (Schrage et al., 2011; OU et al., 2016). Therefore, a systematic review of the performance of existing methods is needed and validations need to be revised if appropriate.

A different example of this is the application of methods to test for skin irritation which were validated for pure chemical substances and mixtures but were of limited use when applied to agrochemical formulations (Kolle et al., 2017). In contrast, reconstructed human cornea-like epithelium models seem to be applicable to assess the eye irritant potential of agrochemical formulations (Kolle et al., 2017). However, applying differing regional classification systems (EU CLP, the United States Environmental Protection Agency (US EPA), The Brazilian Health Regulatory Agency (Anvisa)) to the *in vivo* data as a reference in the validation of *in vitro* tests can affect the performance of *in vitro* methods (Kolle et al., 2012). This illustrates the difficulties imposed on test method developers faced with the need to maintain the validation status of *in vitro* methods, in line with diverse classification requirements (Kolle et al., 2012). Current explored how alternative methods could be used for testing eye irritation hazards of antimicrobial household cleaning products (Curren, 2014). More of such work is needed to adapt and validate alternative methods for testing formulations.

Using alternative methods, skin irritation and corrosion can be addressed by a combination of a skin corrosion and a skin irritation test (OECD TGs 430, 431, 435 and 439) (OECD, 2015d; OECD, 2015e; OECD, 2015f; OECD, 2017b). More complex endpoints, like skin sensitization, would require at least two methods and a data interpretation procedure to combine the results into an overall conclusion on the hazard (Urbisch et al., 2015; Jaworska, 2016). In general, IATAs are used to integrate different test methods and their results. While individual methods have been validated and are covered by OECD's MAD, IATAs are not. There will, however, be no stand-alone alternative methods to address complex toxicological endpoints in the future, but rather comprehensive combinations (Perkins et al., 2015; Rovida et al., 2015; Worth and Patlewicz, 2016). Hence, testing strategies should be validated and IATAs should be accepted in the regulatory arena just as existing animal methods currently are.

Validated and accepted methods and also their combination in

testing strategies and IATAs have limited precision (just like any method). The systematic analysis and consideration of this has only recently started (Hoffmann, 2015; Kolle et al., 2013; Leontaridou et al., 2016; Roberts et al., 2016) and should be extended to all existing and future methods. A test method's precision should be reported in the method description and the resulting uncertainty of all testing results should be clearly described in the study reports. Regulators as well as test developers and users should discuss how precise a method should be for a given application and what level of uncertainty is acceptable. While it may be desirable to have unlimited resources to develop new methods and strategies to assess hazards and risks, it is not realistic. Rather, resources are limited and this has always been reflected, yet not in a systematic way. A systematic approach was recently presented by Gabbert and co-workers using the example of skin sensitization assessment (Gabbert and Weikard, 2013; Gabbert et al., 2017; Leontaridou et al., 2017). The “value of information” provided by a testing method should be used to prioritize and validate methods in the future (Norlen et al., 2014).

2.5. From single methods to defined approaches: ensuring acceptance and use of non-animal approaches

Joao Barroso (JRC)

While the need to conduct animal testing varies around the world, there continues to be progress in terms of reducing, refining or even replacing, to the extent possible, the routine requirement for such studies in decision making processes. Non-animal toxicological methods (e.g. *in vitro* methods, *in silico* predictive models, –omics data) are becoming increasingly available and more relevant from both a biological (i.e., the complexity and physiological significance of the test system) and a mechanistic perspective (in terms of the endpoints measured). In a few cases, *in vitro* test methods have been found, under certain circumstances, to be suitable stand-alone alternatives to animal testing. However, in the majority of cases, data generated with non-animal methods will have to be used in combination to achieve full replacement of *in vivo* testing for the safety assessment of chemicals, because they are not able on their own to cover the variety of endpoints and biological complexity that is usually found in animal tests.

Until recently, a clear international strategy for harmonizing novel data integration approaches that rely on multiple information sources to predict a toxicological endpoint for use in regulatory decisions was missing. Since 2014, however, the OECD started publishing Guidance Documents (GDs) on Integrated Approaches to Testing and Assessment (IATA) for a few human toxicological endpoints, such as skin corrosion/irritation (OECD, 2014a) and serious eye damage/eye irritation (OECD, 2017a), to address this gap. IATA are pragmatic, science-based approaches that rely on an integrated analysis and weighting of relevant existing evidence, from multiple information sources such as physicochemical properties, non-testing methods (e.g., Quantitative Structure-Activity Relationships (QSARs), read-across) and testing methods (e.g., *in vitro*, *in vivo*), and guide the targeted generation of new information where required, to inform regulatory decisions on the hazard identification, hazard characterisation, classification and labelling and/or safety assessment of chemicals. IATA can therefore reduce the overall number of tests that need to be conducted and avoid or reduce animal testing by using *in vivo* methods only as a last resort within the strategy.

Another important concept recently put forward by the OECD in GD No. 255 (OECD, 2016e) is the distinction between IATA and the so-called Defined Approaches (DA). A DA consists of a fixed data interpretation procedure (DIP) applied to data generated with a defined set of information sources. Information sources within DAs typically include *in vitro* methods, physicochemical properties and/or computational methods such as QSARs. The output of a DA can be used either on its own or together with other information sources within an IATA. A generic representation of IATA and its elements, including DAs, is shown in Fig. 4. Whilst IATA are intended to be flexible (e.g., in the

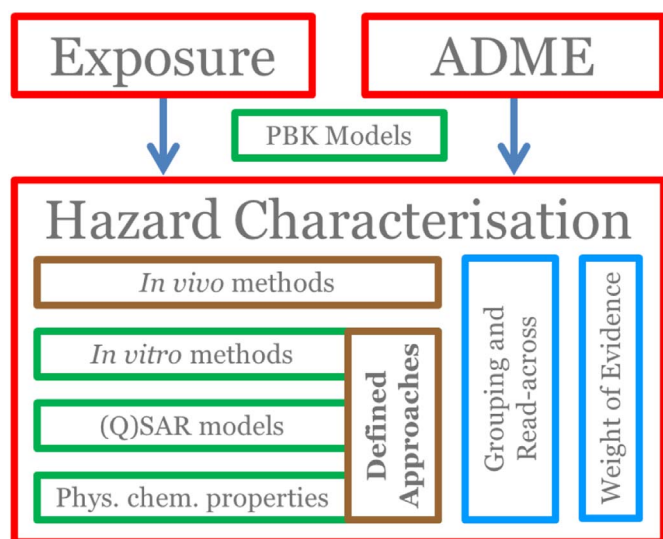


Fig. 4. Generic elements of Integrated Approaches to Testing and Assessment (IATA) and the role of Defined Approaches (DAs) within IATA. DAs standardize the way in which non-animal/mechanistic data are combined to generate outputs comparable to standard *in vivo* test guidelines.

choice of information sources used) and involve a degree of expert judgment (e.g., in the weighting of the information used – IATA assessments are mostly based on weight of evidence), DAs are standardized both in relation to the set of information sources used and in the procedure applied to the data to derive predictions of toxicological effects (i.e., the DIP). The expectation is that (i) the regulatory requirements, (ii) the information sources used in a given chemical assessment, and (iii) the expert judgment applied to those assessments, may vary substantially among regulatory authorities. For this reason, without further prescriptiveness and harmonization, IATA solutions do not easily lend themselves to MAD between authorities in different countries, and hence may lead to an increase in costs for industry and governments performing risk assessments and making classification and labelling decisions. The added value of DAs in this scenario is therefore to provide the necessary standardization in the way information is combined to generate outputs comparable to standard *in vivo* Test Guidelines (TGs), which in turn could lead to further international harmonization in the application of non-animal approaches in the decision-making process and facilitate their acceptance and use.

Skin Corrosion/irritation, serious eye damage/eye irritation and skin sensitization are three areas that have benefited from the development of a considerable number of *in vitro* methods that have been adopted by the OECD, namely TGs 430, 431, 435 and 439 for skin corrosion/irritation (OECD, 2016d; OECD, 2015d; OECD, 2015e; OECD, 2015f), TGs 437, 438, 460, 491 and 492 for serious eye damage/eye irritation (OECD, 2017b; OECD, 2017c; OECD, 2017d; OECD, 2017e; OECD, 2017f), and TGs 442C, 442D and TG 442E for skin sensitization (OECD, 2015a; OECD, 2015g; OECD, 2017g). However, important differences can be observed between these three areas. None of these TGs are full replacement options to the current regulatory *in vivo* tests but, for skin corrosion/irritation, the TGs can all be used stand-alone and a combination of a skin corrosion test (corrosive vs not corrosive) with a skin irritation test (not classified vs classified) offers a full replacement solution (not corrosive + classified = irritant). The risk of dis-harmonization/confusion with several suitable alternatives is limited in this case and therefore, the IATA GD No. 203 on skin corrosion/irritation (OECD, 2014b) guarantees international use and acceptance of *in vitro* data and provides an equal footing between non-animal approaches and the *in vivo* method described in TG 404 (OECD, 2015h). The case of serious eye damage/eye irritation is slightly more complex. In this case, TGs 437, 438, 460 and 491 provide the possibility

to identify chemicals inducing serious eye damage (Cat. 1) but no prediction can be made from all other test results due to considerable rates of Cat. 1 false negatives. TGs 437, 438, 491 and 492 are able to identify chemicals not requiring classification (No Cat.) but again no prediction can be made from all other test results, in this case due to considerable rates of No Cat. false positives. Eye irritation (Cat. 2) can therefore only be concluded from a weight-of-evidence assessment and consequently, the IATA does not provide sufficient certainty on the acceptance of *in vitro* data when “No prediction can be made” results are obtained with the TGs. The development and acceptance of DAs for serious eye damage/eye irritation may be able to close this gap to achieve full replacement of the *in vivo* method described in TG 405 (OECD, 2017h).

Even more complex is the current situation in the area of skin sensitization. Even though, in the (near) future, some *in vitro* methods may alone be able to generate comparable information to the Local Lymph Node assay (LLNA) described in TG 429 (OECD, 2016f) on the skin sensitization potential and potency of chemicals and gain acceptance to be used as full-replacement alternatives, the currently adopted methods, when used in isolation, are not able to generate such information. As stated in the adopted *in vitro* TGs, the generated *in vitro* data should be “considered in the context of an IATA” (i.e. in combination with complementary information) and therefore, the TGs alone do not guarantee the use and acceptance of the *in vitro* methods. To add to this complex scenario, Annex VII to the REACH Regulation was revised in 2016, making the use of *in vitro* methods for skin sensitization testing a standard information requirement and the primary choice over *in vivo* studies. As a mid-term solution, OECD GD No. 255 (OECD, 2016e) was developed and published to provide a set of principles for reporting DAs and templates for their documentation in order to facilitate their assessment and acceptance by regulatory authorities around the world. And in parallel, DAs based on the use of multiple non-animal information sources were developed and documented in Annex 1 to OECD GD No. 256 on the reporting of Defined Approaches to testing and assessment for skin sensitization (OECD, 2016g), using the templates described in OECD GD No. 255. Of note, some of these DAs appear to have comparable or better performance than the LLNA in predicting skin sensitization responses in humans and some are even able to provide information useful for the purpose of skin sensitization potency sub-categorisation (Casati, 2017).

In October 2016, the International Cooperation of Alternative Test Methods (ICATM) organized an international workshop on the regulatory applicability and acceptance of non-animal approaches to skin sensitization, which was hosted by the European Commission's DG Joint Research Centre, to take stock of the latest developments and discuss potential follow-up actions. The workshop was attended by representatives from more than 20 regulatory authorities from the EU, USA, Canada, Japan, S. Korea, Brazil and China. The objectives were: (OECD, 2005) To map the regulatory requirements for skin sensitization across jurisdictions and sectors; (Piersma et al., 2014) To discuss the opportunities and potential obstacles to uptake and acceptance of non-animal methods; (OECD, 2014a) To discuss the evaluation and acceptance processes associated with the use of DAs and IATA; (OECD, 2017a) To discuss potential criteria for the assessment of DAs to facilitate regulatory use, such as e.g., biological plausibility, mechanistic coverage, accessible data interpretation procedures, performance against reference chemicals.

From the discussions it became clear that the certainty required by industry and regulatory authorities on the relevance and international acceptance of the available DAs for skin sensitization to guarantee their deployment and use, may only be achieved if they attain the same level of global acceptance as the standard *in vivo* method(s). In follow-up discussions, the ICATM partners agreed that an alternative approach to skin sensitization assessment that provides equivalent information to the LLNA should be given equivalent regulatory recognition and status. It was also agreed that DAs which combine multiple non-animal

methods can be implemented with equivalent levels of prescriptiveness and reliability in comparison to the LLNA. On this basis, it was concluded that DAs which are shown to be scientifically valid and fit-for-purpose should be incorporated into an OECD TG, as this would enable “equal footing” between those DAs and the LLNA (*i.e.*, it would allow for the replacement of a standard by another standard). A new project proposal was therefore submitted by the EU, USA and Canada to the OECD Test Guidelines Programme in November 2016, proposing the international assessment of DAs for skin sensitization and the development of a Performance-Based Test Guideline (PBTG) for those DAs shown to be valid and to provide a level of information equal to the LLNA. The project was accepted by the Working Group of the National Coordinators for the Test Guidelines Programme (WNT) and is currently underway.

It is important for the future to find the right balance between flexibility and prescriptiveness to embrace new science while allowing for the practicalities of regulatory implementation. IATA facilitate international harmonization while allowing enough flexibility to accommodate scientific developments. However, due to their flexible nature and the consequential enhanced need for expert judgment, it will be more challenging than for DA to have IATA covered by MAD, which requires that data generated in any OECD member country (and non-member adherent to MAD) in accordance with OECD TGs and the Principles of Good Laboratory Practice (GLP) are accepted in other member and adhering countries (see: <http://www.oecd.org/env/ehs/mutualacceptanceofdatamad.htm>). For this reason, IATA descriptions have so far been published in GDs, which may not provide enough certainty in some situations to guarantee use and acceptance of non-animal methods. The standardized nature of DAs, however, should permit their description in TGs, coverage by MAD and increased certainty of acceptance and use.

2.6. Strategic approach to animal-free testing for genotoxicity in case of cosmetic products

Vera Rogiers (SCCS)

Genotoxicity is a key endpoint in the hazard assessment of any chemical as damage to the genetic material might lead to detrimental effects on human health. To assess possible genotoxic actions of chemicals, thus also of cosmetic ingredients, regulatory bodies worldwide recommend, in a first tier, to apply a battery of *in vitro* tests. These tests need to cover the most important genotoxicity endpoints, *i.e.* mutagenicity, clastogenicity and aneugenicity, the latter being structural and numerical changes at the chromosome level. If needed, this is followed up by *in vivo* genotoxicity studies.

By the introduction of an animal testing and marketing ban in the European cosmetic legislation (Regulation 1223/2009/EC), the 3Rs principle (Refinement, Reduction, Replacement) - an important principle that was being gradually introduced into the EU legislation - became for cosmetics restricted to 1R. As a consequence, only validated *in vitro* tests, eventually backed-up by *in chemico* and *in silico* data, had to become the sole predictor of possible genotoxic properties of cosmetic compounds. Using the *in vitro* and *in vivo* genotoxicity data of cosmetic annex substances (colorants, preservatives, UV-filters and hair dyes), available in the different opinions of the Scientific Committee on Consumer Safety (SCCS) issued between 2000 and 2012, important discrepancies were noticed between the *in vitro* and *in vivo* results. Indeed, positive results found through the application of the *in vitro* 3-test battery could most often not be confirmed *in vivo* because of the high sensitivity and very low specificity of the test battery (Ates et al., 2014). Therefore, the 3-test battery was replaced by a 2-test battery composed of an Ames test and an *in vitro* micronucleus test. But still, when one of these tests has a positive outcome, no *in vivo* follow up is allowed and the “positive” compound is lost for cosmetic purposes.

As a possible solution, it was proposed to carry out other *in vitro* tests such as a reconstructed skin micronucleus test or a 3D-based comet

assay, both promising but still under development by the cosmetic industry. Also *in silico* test results and biokinetic data, but in particular molecular and/or mechanistic information, could be used in a Weight of Evidence approach to de-risk a misleading positive *in vitro* result. In that context the PhD research of Gamze Ates was presented in which a testing strategy was developed which integrates the current 2-test *in vitro* genotoxicity battery either with computational tools (Ates et al., 2016), reporter-based assays (Ates et al., 2016) or genomic fingerprinting (Ates, 2017) to de-risk a misleading positive *in vitro* genotoxicity test. The results can be summarized as follows: (i) By integrating a variety of *in silico* tools (SARpy, TOXTREE, CAESAR, Derek, T.E.S.T., *IN SILICO* BATTERY) a sensitivity of 53–87% and a specificity of 52–71% could be obtained; (ii) By integrating 2 reporter-based assays (ToxTracker and Vitotox) a sensitivity of 93% and specificity of 73% was found, which shows a better performance than using only an Ames test and an *in vitro* micronucleus test; (iii) The best results were obtained by integrating genomic fingerprinting, which was realized by developing a qPCR-array, based on an 84-gene fingerprint. The latter could be retrieved from gene expression analyses on whole genome microarrays, using the human HepaRG cell line as *in vitro* model. Testing the qPCR array with 12 known genotoxic and non-genotoxic compounds showed a sensitivity and specificity of 83%. Other genomic fingerprint data existing in this field use either human TK 6 cells or HepG2 cells, which have a more limited biotransformation capacity. Consequently, obtaining mechanistic/molecular information by using a dedicated qPCR array based upon human HepaRG cells exposed to the compound under consideration, is proposed as a strategic and pragmatic approach, which together with expert advice, seems to provide an improved way forward for cosmetic ingredients in case of misleading positive results when no *in vivo* testing is possible. Such mechanistic approaches add weight to the evidence.

3. Breakout discussion group summaries

3.1. Breakout group 1: how can the quality of alternative animal-free methods be improved?

Chantra Eskes (SeCAM) and Michael Oelgeschläger (BfR)

The breakout group addressed a number of key questions in relation to quality control of single methods and testing batteries. There was general agreement that in the future a comprehensive assessment of a single method prior or after validation will be less pertinent, since most likely testing strategies and DAs composed of combinations of test methods will be needed. Any test method used within a testing strategy or DA should always be robust and provide an added value with respect to (regional) availability or applicability. In addition, the method's relevance should be evaluated in terms of its mechanistic added value and contribution to the final predictive capacity of a testing strategy. It was considered particularly important that the same set of chemicals are assessed in the various test methods forming a test strategy or DA that target an adverse effect, in order to facilitate the development of DIP. The problem of testing concentrations of chemicals to the limit of solubility was acknowledged but reliable exposure data that would allow the definition of a relevant upper concentration limit is scarce and, thus, more work needs to be done to address this issue. There are increasing numbers of methods available that have not undergone formal validation but might still be valuable for *in vitro* testing as part of DAs. For example, in case of high-throughput screening techniques, the following elements were found important to warrant acceptance of the data: use of a battery of internal controls and reference chemicals, the coverage of biologically plausible mechanisms, the evaluation of within-laboratory reproducibility and a thorough analysis of primary data in order to assess the plausibility of data interpretation.

In general, a broadening of the applicability domain of *in vitro* methods to mixtures was also identified as an important issue that needs to be addressed. Mechanistic-based *in vitro* approaches can work

for mixtures, although for certain types of mixtures or formulations, single (and potentially unknown) components might interfere with the test system. However, it was acknowledged that the definition of mixtures is difficult to address as clear and harmonized definitions are hardly achievable. Similar problems exist in the area of pharmaceutical products where drug interactions are also difficult to predict. In any case, the applicability of a test method needs to be considered taking into account mechanistic aspects. Feedback from users on a regular basis (e.g., every 4–5 years) might help to develop a better understanding of the applicability of test methods for various types of mixtures as well as chemicals over time that could also be used to update and improve already existing test guidelines.

Altogether, it was agreed that a comprehensive description of the biological system is of utmost importance, since the relevance or predictivity of a method depends on its ability to mimic (human) physiological processes *in vivo*. Following this line, *in vitro* data should be compared with the human situation in terms of physiology and toxicological understanding rather than with experimental animal data. This approach obviously depends on sufficient knowledge of the molecular mechanism(s) mediating adverse effects and the development and use of AOP was considered helpful, although this will not always be possible. The use of human-based *in vitro* systems might allow a higher predictivity than the current animal models. However, to confirm this, one would need reliable human data, which is rarely available, and closer cooperation with clinicians and biomedical researches was proposed to address this issue. There was general agreement that compound potency and toxicokinetic aspects have to be considered and the further development of suitable (human) PBTK models that in combination with reverse dosimetry might facilitate quantitative *in vitro*-*in vivo* extrapolation (QIVIVE) were generally regarded of central importance. All these various aspects of method description and use depend on a clear definition of an acceptable level of uncertainty. A level of uncertainty should be defined for all components of a testing strategy and DA. However, it also became clear that the way how “uncertainty” of a given approach is determined is not defined yet and might relate to its applicability domain, species differences as well as the level of mechanistic coverage. The EFSA guideline on uncertainty was considered a good starting point and AOPs were considered as a useful tool to evaluate at least the level of confidence in a defined biological process.

Finally, the current description of requirements and general guidance to conduct *in vitro* methods under GLP might be insufficient to ensure that test methods that have already received regulatory acceptance, e.g. as OECD test guidelines, are performed appropriately. However, the current development of the Guidance document on Good *In Vitro* Method Practices (GIVIMP) might at least partially solve this problem. In addition, the usual quality control measures according to ISO 17025, including small ring trials with blinded compounds, allows to ensure laboratory proficiency. One problem that was identified is the lack of sufficient detail in OECD guidelines that always need to be complemented by SOPs but often also require additional guidance. In particular, critical steps in the procedure are not explicitly described, but would be helpful to both test method users and quality auditors. Thus, it was proposed to describe critical steps already in the respective test guidelines.

3.2. Breakout group 2: “How can the validation process be accelerated?”

Maurice Whelan (JRC) and Anne Kienhuis (RIVM)

This breakout group addressed acceleration of the validation process and discussed factors that need to change to increase efficiency and effectiveness of validation studies. The breakout group agreed that instead of investing in acceleration of the classical approach to validation (e.g. *in vitro* methods subject to ring-trials) the validation framework should be changed rigorously, which would require a paradigm shift in the regulatory landscape. It was acknowledged that this is complex and requires a long term investment.

A limiting aspect in the current approach to validation and acceptance of alternative methods that was discussed in the breakout group was the reference to the animal study as the gold standard. It was mentioned that it is sometimes perceived that regulatory agencies are reluctant to talk about the limitations of animal data. It is also experienced that more is asked for validation of alternative methods than for “gold standard” *in vivo* methods. A suggested solution may be the use of appropriate benchmarks that are compared to the human situation instead of using one-in-one comparison of alternatives with animal tests. It was addressed that, compared to *in vivo* studies, non-animal alternatives produce new types of data, which on the one hand means that some information may be lost but on the other hand, new information may be gained. To gain as much information as possible from non-animal alternatives, human relevant and mechanism-based testing strategies instead of individual methods or tools should be used. Since focusing on limitations restricts thinking, the break out group suggested focusing on what works, even if it results in accepting a certain level of uncertainty, which is already the case today when using the so-called “gold standard” *in vivo* methods.

It was observed that the validation in its current form is perceived as non-attractive, considering complexity, costs and time. A new validation framework should start with the establishment of the essential elements that are required in the new framework. These include biological relevance, transferability and reproducibility. With the new technologies evidence should be collected that is covering the biological domain. Furthermore, the methods need to be used by many people and should generate similar results. When transferability and reproducibility are not properly organized, it can affect trust in the system. Based on these principles, it can be discussed how new technologies, such as omics, can be considered in the regulatory framework. Otherwise, technologies that are originally developed to accelerate the process might instead decelerate progress. Furthermore, validation can be made more attractive for scientists in academia. A suggestion was to change the term “validation” into “quality assurance” and to integrate the process in grant programs. Resources and finances are needed to keep evaluating and improving the new technological advancements. Next to the investment in advanced and complex technologies, it should be taken into account that there might be modest, low tech solutions as well.

In the current validation framework, the OECD is a democratic organization with a large number of test guidelines and projects on the work plan, working with diplomacy and consensus, and dealing with many stakeholders in the same strategic level playing field. This should be acknowledged since a rigorous change of this system may affect socioeconomic legal implications. Validated methods and OECD guidelines are a legal obligation to obtain H (hazard) sentences for CLP (Classification, Labeling and Packaging of chemical substances and mixtures). It was suggested in the breakout group to use the democracy represented by OECD in the current framework as a driver to accelerate the validation process. It is observed that within the big machinery of the OECD, new methods that are critically needed are validated quite fast. In these cases, there is a sense of intrinsic priority to get democracy in action. On the one hand, the framework needs a certain degree of prescription. On the other hand, the acceptance of new methods requires some flexibility within the framework.

In the concluding part of the group discussion, it was proposed that a global organization, a strategic vision, and an integrative approach are needed. Specifically, different stakeholders (academia, clinicians, toxicologists, regulators, industry) need to work closer as a team, already at the start of the development of a new fit for purpose method or strategy. Although conflicts of interest will exist, stakeholders should focus on common interests and decide what needs to be done. In the end, all stakeholders aim to provide the best, safe products on the market, when possible, avoiding using animals. Key to a successful validation framework is collaboration and communication which should be coordinated, fluent and easy. The break out group suggested

that the efficient development, validation and acceptance of methods can be achieved by establishing a multi-stakeholder working group. Together, they can define common goals, identify hypothesis, address endpoints, select chemicals and assess quality of methods. These multi-stakeholder working groups can work on so-called case studies. It was concluded that these case studies could serve as the new validation (or 'quality assurance').

3.3. Breakout group 3: "How can regulatory animal-free approaches be facilitated?"

Jan van Benthem (RIVM) and Janine Ezendam (RIVM)

First, the group identified main issues that have an impact on regulatory acceptance and implementation of novel approaches. One is that regulators prefer methods that provide them with the same level of information and/or read-out as the current (animal) standard. Modern toxicology assays, however, may deliver different read-outs and *in vitro* assays do not always give a quantitative measure. Those assays that do provide dose-response data may be used to derive a no-effect level, but it is not yet fully understood how well those correlate with a DNEL or NO(A)EL which are essential to adequately perform a safety assessment. Quantitative *in vitro* – *in vivo* extrapolation (QIVIVE) modelling is evolving, which will lead to further understanding whether *in vitro* quantitative data can be used to derive an equivalent DNEL/NO(A)EL. Furthermore, the current paradigm does not allow easy introduction of innovative approaches that mostly measure mechanisms of toxicity rather than an apical endpoint. Clearly, a paradigm shift is needed to allow incorporation of novel non-animal approaches, although it was acknowledged that this is a far from simple task.

An important observation was that regulators have different roles and influence in the process of regulatory acceptance and implementation. Some regulators are involved early, whereas others become involved at the end of the process or not at all. Lack of familiarity with novel approaches can make a regulator reluctant to approve its implementation. Some regulators are already involved in the early evaluation of the regulatory relevance of a new method (e.g. EURL ECVAM's PARERE network) or at the OECD level in the test guideline program. Many regulators are only performing dossier evaluations, often with time constraints. They have insufficient time to become acquainted with novel non-animal approaches and keep pace with the scientific progress made in this area. For these reasons, they will be cautious to use these new approaches. In addition, test method developers often have insufficient insight into the needs and thoughts of regulators. These observations indicate the need to encourage and improve interactions between test developers and regulators/risk assessors. Also, due to the different vertical legislations for chemical safety there are multiple enforcement bodies that play different roles in the process towards regulatory acceptance and implementation. Hence, there is an urgent need for harmonization and to share best practices and experiences to facilitate regulatory acceptance of novel methods in all relevant regulations. Early involvement of regulators in the process of development of new methods is therefore considered of importance, but may be challenging. Experience from EURL ECVAM with their PARERE network shows that often regulators find it difficult to evaluate upfront whether a method that is not yet validated is fit-for-regulatory purposes.

Another factor that was considered to hamper the implementation of 3R approaches was the United Nations Globally Harmonized System of Classification and Labelling of Chemicals (GHS) (UN, 2017) and the European Union Regulation on Classification, Labelling and Packaging of chemicals (CLP) implementing UN GHS in EU (EC, 2008). GHS/CLP is an important system to inform users about chemical hazards through the use of standard symbols and terms. Information on hazards is obtained according to GHS/CLP criteria, which in its current form does not allow the use of *in vitro* data. Since CLP criteria are important in several downstream regulations, some discussants considered this an

important hurdle in the acceptance of non-animal approaches. Hence, targeted actions to modify these criteria are essential to ensure a broader use of the validated and relevant non-animal methods. One important initiative is undertaken by the Netherlands, trying to adapt the GHS criteria allowing the use of *in vitro* data to assess skin corrosion/irritation properties.

The group discussed as the third point the choice of the benchmark ('gold standard') used to compare the predictivity of the new method or testing strategy. Traditionally data from animal tests are used as a benchmark, but the concern was expressed that uncertainties and variability of these reference data are not fully understood nor are they currently being considered during the process of validation and acceptance of alternative approaches. Another drawback of the sole reliance on animal reference data is that for many "novel" human-relevant toxicological endpoints no or limited animal data are available, e.g. developmental neurotoxicity and endocrine disruption. Clearly, there is a need for another benchmark that informs on the mechanistic relevance for a specific endpoint. Understanding of human biology or a clear mechanistic description of a toxicological endpoint in an AOP or MoA was considered essential for this purpose.

The fourth topic on the optimal way to design and validate a testing strategy in such a way that it satisfies regulatory requirements or needs was also shortly discussed. No solutions or recommendations were made during this workshop, but some valid points were raised. The modular approach used for an individual test method (Hartung et al., 2004) is applicable to testing strategies as well, although the relative importance of the different modules and the order under which they should be addressed may need to be different when moving from the validation of individual methods to validation of testing strategies or of batteries of tests. Regulators prefer a DA that consists of a predefined number of assays, a fixed prediction model and a clear-cut outcome, over an IATA. An IATA offers more flexibility in selection of the methods performed for a specific chemical and the decision is based on a weight-of-evidence approach, leaving more room for subjective interpretation. From a scientific perspective IATA has many advantages, because it can be a targeted chemical-specific approach that allows selection of the most appropriate methods and incorporation of novel methods when they become available. From the perspective of a regulator, however, IATA may be more difficult to accept, because of the flexibility in both test method selection as well as data interpretation.

The last topic discussed in this group was how DAs and eventually also non-animal IATA solutions can be on equal footing with individual *in vivo* test guidelines, for which MAD applies. It was anticipated by the group that it would be easier to develop a test guideline for a DA, than for an IATA. More clear guidance is needed on how to perform a weight-of-evidence approach in order to standardize IATA-based assessments to a level permitting MAD. The downside is that once novel individual assays become available, it will be difficult to adapt the DA. The scope, limitations and uncertainties of the DA need to be clearly defined. Furthermore, an easy prediction model or DIP is preferred by regulators.

In summary, Europe needs a taskforce that coordinates knowledge and data sharing between different international regulatory bodies and EU member states to enable harmonized use of novel non-animal approaches. It is important to characterize upfront what the method is fit for and be clear and transparent about its limitations. Communication between different stakeholders (academia, regulators, industry) should be improved and we should invest in clear terminology and guidance.

4. Outlook

After decades of dedicated research into animal-free methods for chemical hazard and risk assessment, we are still pondering the question of how validation, acceptance and regulatory implementation of these methods can be accelerated. We do note that much important work has been done on the development of alternative methods. Many

in vitro methods are proving useful at least in a screening environment, e.g. to provide first indications of compound effects and for prioritization purposes among chemical analogs. However, replacement of animal studies in regulatory settings has been very scarce. This stagnation was mainly caused by the use of the animal study as the gold standard for human hazard assessment, and the consequential procedure to validate an alternative assay one to one against the animal study data. The reductionist nature of *in vitro* alternatives in most cases limits their individual validity for predicting toxicity in the intact organism. In addition, superficial *in vitro* to *in vivo* comparisons have reduced the impact of validation studies, such as scoring compounds as positive or negative, whereas it is the potency that determines toxicity.

The present workshop has pointed out that the classical bottom-up development and validation of alternatives for regulatory toxicology has to be balanced by a top-down approach from the perspective of regulatory needs based on scientific knowledge. Moreover, these needs should not be distilled from current regulatory practice, as it is largely animal-centered. It should begin with the question which parts of human biology needs to be covered in test systems in order to be able to sufficiently detect all toxicities. Downstream follows the design of testing strategies and the selection of a series of appropriate complementary assays for critical key events covering the network of toxicity pathways. This top-down approach is needed to convince scientists and regulators that the necessary biology is comprehensively incorporated into the novel animal-free system. Current activities to design ontology-driven computational systems toxicology approaches as well as IATA and DA show practical applications of this principle. Validation of these integrated approaches in terms of predictability is needed at the level of the strategy, and not at the level of the individual assay. In that respect, it is often considered that case studies at the level of the strategy are the new validation. At the same time the bottom-up aspect remains important, as definition of biological and chemical space and technical validation (e.g. reproducibility and transferability) of individual assays is crucial for determining the usefulness and positioning of individual assays within a testing strategy.

The biology-driven testing-strategy-focused approach of developing alternative approaches towards human chemical hazard and risk assessment is still rather revolutionary. It does not build on current regulatory requirements but starts from the question of the biology that needs to be covered. This naturally meets with reservations from various stakeholders in the current regulatory arena. In the present workshop, experts from a variety of stakeholders including academia, industry, non-governmental organizations and regulatory agencies discussed this innovative approach with an open mind. An open mindset of all stakeholders and active communication between developers and regulators throughout the process are prerequisites for providing a fundament for a transition towards human chemical safety assessment that is mechanistically driven and that avoids the detour of the animal experiment. If anything, this workshop has proven an animated multi-stakeholder podium for developing and sharing ideas supporting novel avenues for chemical risk assessment for man. The challenge lies in keeping the momentum going, and to uphold the necessary pace in scientific progress to provide continuous justification for the optimism as regards the feasibility of these approaches.

Transparency document

The [Transparency document](#) associated with this article can be found, in online version.

Acknowledgments

The authors would like to acknowledge the contributions of the remaining workshop participants:

J. Coenen (SET), S. Escher (ITEM), N. Gellatly (NC3Rs), A. Gourmelon (OECD), M. Hermann (Henkel/IKW), M. Herzler (BfR), T.

Höfer (BfR), M. Peiser (BfR), D. Petersohn (Henkel), J. Scheel (Beiersdorf), G. Schönfelder (BfR), M. Timmermann (Science Europe), T. Tralau (BfR).

References

- Ankley, G.T., Bennett, R.S., Erickson, R.J., Hoff, D.J., Hornung, M.W., Johnson, R.D., et al., 2010. Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. *Environ. Toxicol. Chem.* 29 (3), 730–741.
- Ates, G., 2017. Thesis to Obtain the Degree of Doctor in Pharmaceutical Sciences. Free University Brussels, Brussels.
- Ates, G., Doktorova, T.Y., Pauwels, M., Rogiers, V., 2014. Retrospective analysis of the mutagenicity/genotoxicity data of the cosmetic ingredients present on the Annexes of the Cosmetic EU legislation (2000–12). *Mutagenesis* 29 (2), 115–121.
- Ates, G., Favys, D., Hendriks, G., Derr, R., Mertens, B., Verschaeve, L., et al., 2016. The Vitotox and ToxTracker assays: a two-test combination for quick and reliable assessment of genotoxic hazards. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*. 810, 13–21.
- Ates, G., Raitano, G., Heymans, A., Van Bossuyt, M., Vanparys, P., Mertens, B., et al., 2016. In silico tools and transcriptomics analyses in the mutagenicity assessment of cosmetic ingredients: a proof-of-principle on how to add weight to the evidence. *Mutagenesis* 31 (4), 453–461.
- Basketter, D., Kolle, S.N., Schrage, A., Honarvar, N., Gamer, A.O., van Ravenzwaay, B., et al., 2012. Experience with local lymph node assay performance standards using standard radioactivity and nonradioactive cell count measurements. *J. Appl. Toxicol.* 32 (8), 590–596.
- Burden, N.M.C., Müller, B.P., Terry, C., Westmoreland, C., Kimber, I., 2015. Aligning the 3Rs with new paradigms in the safety assessment of chemicals. *Toxicol. in Vitro* 330, 62–66.
- Campo-Paysaa, F., Marlétaz, F., Laudet, V., Schubert, M., 2008. Retinoic acid signaling in development: tissue-specific functions and evolutionary origins. *Genesis* 46 (11), 640–656.
- Casati, S., 2017. Contact hypersensitivity: integrated approaches to testing and assessment. *Current Opinion in Toxicology* 5, 1–5.
- Commission E. Regulation (EC), 2008. No 1272/2008 of the European Parliament and of the Council of 16 December 2008 on Classification, Labelling and Packaging of Substances and Mixtures. EUR-Lex.
- Commission E. Regulation (EC), 2006. No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH). EUR-Lex. (Document 02006R1907-20140410).
- Committee EESA, 2016. ESAC opinion on the use of performance standards to evaluate test methods similar to a validated reference method. In: ESAC Opinion No 2016–06 of 24 June 2016; EUR 28183 EN.
- Curren, R., 2014. Cooperation Between the United States Environmental Protection Agency and Industry to Develop an In Vitro Ocular Hazard Testing Strategy. Science and the Law: Analytical Data in Support of Regulation in Health, Food, and the Environment. pp. 111–118.
- EC, 2008. Regulation (EC) No 1272/2008 of the European parliament and of the council of 16 December 2008 on classification, labelling and packaging of substances and mixtures, amending and repealing directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) No 1907/2006. Off. J. Eur. Union L353, 1–1355.
- Edwards, S.W., Tan, Y.-M., Villeneuve, D.L., Meek, M.E., McQueen, C.A., 2015. Adverse outcome pathways—organizing toxicological information to improve decision making. *J. Pharmacol. Exp. Ther.* 356 (1), 170.
- Gabbert, S., Weikard, H.-P., 2013. Sequential testing of chemicals when costs matter: a value of information approach. *Hum. Ecol. Risk Assess.* 19, 1067–1088.
- Gabbert, S., Leontaridou, M., Landsiedel, R., 2017. A critical review of adverse outcome pathway-based concepts and tools for integrating information from non-animal testing methods: the case of skin sensitization. *Applied In Vitro Toxicology* 3 (3), 250–264. <http://dx.doi.org/10.1089/aivt.2017.0015>.
- Hartung, T., Bremer, S., Casati, S., Coecke, S., Corvi, R., Fortaner, S., et al., 2004. A modular approach to the ECVAM principles on test validity. *Alternatives to Laboratory Animals: ATLA*. 32 (5), 467–472.
- Hoffmann, S., 2015. LLNA variability: an essential ingredient for a comprehensive assessment of non-animal skin sensitization test methods and strategies. *ALTEX* 32 (4), 379–383.
- Hutson, M.S., Leung, M.C.K., Baker, N.C., Spencer, R.M., Knudsen, T.B., 2017. Computational model of secondary palate fusion and disruption. *Chem. Res. Toxicol.* 30 (4), 965–979.
- Jaworska, J., 2016. Integrated testing strategies for skin sensitization hazard and potency assessment—state of the art and challenges. *Cosmetics* 3, 16.
- Kleinstreuer, N., Dix, D., Rountree, M., Baker, N., Sipes, N., Reif, D., et al., 2013. A computational model predicting disruption of blood vessel development. *PLoS Comput. Biol.* 9 (4), e1002996.
- Kolle, S.N., Basketter, D.A., Casati, S., Stokes, W.S., Strickland, J., van Ravenzwaay, B., et al., 2013. Performance standards and alternative assays: practical insights from skin sensitization. *Regul. Toxicol. Pharmacol.* 65 (2), 278–285.
- Kolle, S.N., Sullivan, K.M., Mehling, A., van Ravenzwaay, B., Landsiedel, R., 2012. Applicability of in vitro tests for skin irritation and corrosion to regulatory classification schemes: substantiating test strategies with data from routine studies. *Regulatory Toxicology and Pharmacology: RTP*. 64 (3), 402–414.
- Kolle, S.N., Van Cott, A., van Ravenzwaay, B., Landsiedel, R., 2017. Lacking applicability of in vitro eye irritation methods to identify seriously eye irritating agrochemical

- formulations: results of bovine cornea opacity and permeability assay, isolated chicken eye test and the EpiOcular™ ET-50 method to classify according to UN GHS. *Regul. Toxicol. Pharmacol.* 85, 33–47.
- Kolle, S.N., van Ravenzwaay, B., Landsiedel, R., 2017. Regulatory accepted but out of domain: in vitro skin irritation tests for agrochemical formulations. *Regul. Toxicol. Pharmacol.* 89 (Supplement C), 125–130.
- Landsiedel, R., Kolle, S.N., Bauch, C., 2013. Entwicklung und Implementierung einer tierversuchsfreien Strategie zur Prüfung der lokalen Toxizität von Chemikalien. Verleihung des 32. Tierschutz-Forschungspreises 2013 des Bundesministeriums für Ernährung, Landwirtschaft und Verbraucherschutz (BMELV) durch den Parlamentarischen Staatssekretär Peter Bleser. <http://dx.doi.org/10.5281/zenodo.1186984>.
- Leontaridou, M., Gabbert, S., van Ierland, E.C., Worth, A.P., Landsiedel, R., 2016. Evaluation of non-animal methods for assessing skin sensitisation hazard: a Bayesian Value-of-Information analysis. *Altern. Lab. Anim.* 44 (3), 255–269.
- Leontaridou, M., Urbisch, D., Kolle, S.N., Ott, K., Mulliner, D.S., Gabbert, S., Landsiedel, R., 2017. The borderline range of toxicological methods: quantification and implications for evaluating precision. *ALTEX* 34 (4), 525–538.
- Leung, M.C.K., Hutson, M.S., Seifert, A.W., Spencer, R.M., Knudsen, T.B., 2016. Computational modeling and simulation of genital tubercle development. *Reprod. Toxicol.* 64, 151–161.
- Norlen, H., Worth, A.P., Gabbert, S., 2014. A tutorial for analysing the cost-effectiveness of alternative methods for assessing chemical toxicity: the case of acute oral toxicity prediction. *ATLA* 42 (2), 115–127.
- OECD, 2005. Guidance Document No. 34 on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment. OECD Environment, Health and Safety Publications. Series on Testing and Assessment.
- OECD, 2010a. OECD Test Guideline No. 429 Skin Sensitization: Local Lymph Node Assay. OECD Series on Testing and Assessment.
- OECD, 2010b. OECD Test Guideline No. 442A. Skin Sensitization: Local Lymph Node Assay: DA OECD Series on Testing and Assessment.
- OECD, 2010c. OECD Test Guideline No. 442B. Skin Sensitization: Local Lymph Node Assay: BrdU-ELISA OECD Series on Testing and Assessment.
- OECD, 2012a. The adverse outcome pathway for skin sensitisation initiated by covalent binding for proteins. Part 1. Scientific evidence. OECD Environment, Health and Safety Publications Series on Testing and Assessment 168, 1–46.
- OECD, 2012b. The adverse outcome pathway for skin sensitisation initiated by covalent binding for proteins. Part 2. Use of the AOP to develop chemical categories and integrated assessment and testing approaches. OECD Environment, Health and Safety Publications Series on Testing and Assessment 168, 47–80.
- OECD, 2014a. New Guidance Document on an Integrated Approach on Testing and Assessment (IATA) for Skin Corrosion and Irritation. OECD Series on Testing and Assessment. (No. 203).
- OECD, 2014b. Guidance Document on an Integrated Approach on Testing and Assessment (IATA) for Skin Corrosion and Irritation. Series on Testing and Assessment (No. 203).
- OECD, 2015a. OECD Test Guidelines for the Testing of Chemicals. Section 4: Health Effects, TG 442C In Chemico Skin Sensitisation: Direct Peptide Reactivity Assay (DPRA).
- OECD, 2015b. OECD Test Guidelines for the Testing of Chemicals. Section 4: Health Effects, TG 442D In Vitro Skin Sensitisation: ARE-Nrf2 Luciferase Test Method.
- OECD, 2015c. Test No. 439: In Vitro Skin Irritation - Reconstructed Human Epidermis Test Method.
- OECD, 2015d. In Vitro Skin Corrosion: Transcutaneous Electrical Resistance Test Method (TER). OECD Guideline for the Testing of Chemicals No. 430.
- OECD, 2015e. In Vitro Membrane Barrier Test Method for Skin Corrosion. OECD Guideline for the Testing of Chemicals No. 435.
- OECD, 2015f. In Vitro Skin Irritation: Reconstructed Human Epidermis Test Method. OECD Guideline for the Testing of Chemicals No. 439.
- OECD, 2015g. In Vitro Skin Sensitisation: ARE-Nrf2 Luciferase Test Method. OECD Guideline for the Testing of Chemicals No. 442D.
- OECD, 2015h. Acute Dermal Irritation/Corrosion. OECD Guideline for Testing of Chemicals No. 404.
- OECD, 2016a. OECD Test Guidelines for the Testing of Chemicals. Section 4: Health Effects, TG 442E In Vitro Skin Sensitisation: Human Cell Line Activation Test (h-CLAT).
- OECD, 2016b. Guidance Document for the Use of Adverse Outcome Pathways in Developing Integrated Approaches to Testing and Assessment (IATA) Series on Testing & Assessment. (No. 260).
- OECD, 2016c. Guidance Document on the Reporting of Defined Approaches and Individual Information Sources to be Used Within Integrated Approaches to Testing and Assessment (IATA) for Skin Sensitisation Series on Testing & Assessment. (No. 256).
- OECD, 2016d. In Vitro Skin Corrosion: Reconstructed Human Epidermis (RHE) Test Method. OECD Guideline for the Testing of Chemicals No. 431.
- OECD, 2016e. Guidance Document on the Reporting of Defined Approaches to be Used Within Integrated Approaches to Testing and Assessment. Series on Testing and Assessment (No. 255).
- OECD, 2016f. Guidance Document on the Reporting of Defined Approaches to be Used Within Integrated Approaches to Testing and Assessment Series on Testing & Assessment. (No. 255).
- OECD, 2016g. Guidance Document on the Reporting of Defined Approaches and Individual Information Sources to be Used Within Integrated Approaches to Testing and Assessment (IATA) for Skin Sensitisation. Series on Testing and Assessment No. 256.
- OECD, 2017a. Guidance Document on an Integrated Approach on Testing and Assessment (IATA) for Serious Eye Damage and Eye Irritation. Series on Testing and Assessment.
- OECD, 2017b. Bovine Corneal Opacity and Permeability Test Method for Identifying (i) Chemicals Inducing Serious Eye Damage and (ii) Chemicals Not Requiring Classification for Eye Irritation or Serious Eye Damage. OECD Guideline for Testing of Chemicals No. 437.
- OECD, 2017c. Isolated Chicken Eye Test Method for Identifying (i) Chemicals Inducing Serious Eye Damage and (ii) Chemicals Not Requiring Classification for Eye Irritation or Serious Eye Damage. OECD Guideline for Testing of Chemicals No. 438.
- OECD, 2017d. Fluorescein Leakage Test Method for Identifying Ocular Corrosives and Severe Irritants. OECD Guideline for Testing of Chemicals No. 460.
- OECD, 2017e. Short Time Exposure In Vitro Test Method for Identifying (i) Chemicals Inducing Serious Eye Damage and (ii) Chemicals Not Requiring Classification for Eye Irritation or Serious Eye Damage. OECD Guideline for the Testing of Chemicals No. 491.
- OECD, 2017f. Reconstructed Human Cornea-like Epithelium (RhCE) Test Method for Identifying Chemicals not Requiring Classification and Labelling for Eye Irritation or Serious Eye Damage. OECD Guideline for the Testing of Chemicals No. 492.
- OECD, 2017g. Key Event Based Test Guideline: In Vitro Skin Sensitisation Assays Addressing the AOP Key Event on Activation of Dendritic Cells. OECD Guideline for the Testing of Chemicals No. 442E.
- OECD, 2017h. Acute Eye Irritation/Corrosion. OECD Guideline for the Testing of Chemicals No. 405.
- OU, Buesen R., Sauer, U.G., Landsiedel, R., 2016. Acute oral toxicity testing: scientific evidence and practicability should govern Three Rs activities. *ATLA* 44 (4), 391–398.
- Perkins, E.J., Antczak, P., Burgoon, L., Falciani, F., Garcia-Reyer, N., Gutsell, S., et al., 2015. Adverse outcome pathways for regulatory applications: examination of four case studies with different degrees of completeness and scientific confidence. *Toxicol. Sci.* 148 (1), 14–25.
- Piersma, A.H., Ezendam, J., Luijten, M., Muller, J.J.A., Rorije, E., van der Ven, L.T.M., et al., 2014. A critical appraisal of the process of regulatory implementation of novel in vivo and in vitro methods for chemical hazard and risk assessment. *Crit. Rev. Toxicol.* 24 (0), 1–19.
- Ramirez, T., Beken, S., Chlebus, M., Ellis, G., Griesinger, C., De Jonghe, S., et al., 2015. Knowledge sharing to facilitate regulatory decision-making in regard to alternatives to animal testing: report of an EPAA workshop. *Regul. Toxicol. Pharmacol.* 73 (1), 210–226.
- Richard, A.M., Judson, R.S., Houck, K.A., Grulke, C.M., Volarath, P., Thillainadarajah, I., et al., 2016. ToxCast chemical landscape: paving the road to 21st century toxicology. *Chem. Res. Toxicol.* 29 (8), 1225–1251.
- Roberts, D.W., Schultz, T.W., Api, A.M., 2016. Chemical applicability domain of the Local Lymph Node Assay (LLNA) for skin sensitisation potency. Part 3. Apparent discrepancies between LLNA and GPMT sensitisation potential: false positives or differences in sensitivity? *Regulatory Toxicology and Pharmacology: RTP* 80, 260–267.
- Rovida, C., Alepee, N., Api, A.M., Basketter, D.A., Bois, F.Y., Caloni, F., et al., 2015. Integrated testing strategies (ITS) for safety assessment. *ALTEX* 32 (1), 25–40.
- Schrage, A., Hempel, K., Schulz, M., Kolle, S.N., van Ravenzwaay, B., Landsiedel, R., 2011. Refinement and reduction of acute oral toxicity testing: a critical review of the use of cytotoxicity data. *Alternatives to Laboratory Animals: ATLA* 39 (3), 273–295.
- Schrage, A., Kolle, S.N., Moreno, M.C., Norman, K., Raabe, H., Curren, R., et al., 2011. The bovine corneal opacity and permeability test in routine ocular irritation testing and its improvement within the limits of OECD test guideline 437. *Alternatives to Laboratory Animals: ATLA* 39 (1), 37–53.
- Tollefsen, K.E., Scholz, S., Cronin, M.T., Edwards, S.W., de Knecht, J., Crofton, K., et al., 2014. Applying adverse outcome pathways (AOPs) to support integrated approaches to testing and assessment (IATA). *Regul. Toxicol. Pharmacol.* 70 (3), 629–640.
- Tonk, E.C.M., Pennings, J.L.A., Piersma, A.H., 2015. An adverse outcome pathway framework for neural tube and axial defects mediated by modulation of retinoic acid homeostasis. *Reprod. Toxicol.* 55, 104–113.
- UN, 2017. United Nations globally harmonized system of classification and labelling of chemicals (GHS). New York, USA and Geneva, Switzerland. (Contract No.: ST/SG/AC.10/30/Rev.7).
- Urbisch, D., Mehling, A., Guth, K., Ramirez, T., Honarvar, N., Kolle, S., et al., 2015. Assessing skin sensitization hazard in mice and men using non-animal test methods. *Regul. Toxicol. Pharmacol.* 71 (2), 337–351.
- USA NNAoS, 2007. Toxicity Testing in the Twenty-First Century: A Vision and a Strategy. National Academies Press, Washington, DC.
- Worth, A.P., Patlewicz, G., 2016. Integrated approaches to testing and assessment. In: Eskes, C., Whelan, M. (Eds.), *Validation of Alternative Methods for Toxicity Testing*. Springer International Publishing, Cham, pp. 317–342.
- Wu, S., Fisher, J., Naciff, J., Laufersweiler, M., Lester, C., Daston, G., et al., 2013. Framework for identifying chemicals with structural features associated with the potential to act as developmental or reproductive toxicants. *Chem. Res. Toxicol.* 26 (12), 1840–1861.